Eur J Clin Chem Clin Biochem 1996; 34:423–430

Error Models and Quality Control Performance

Andrea Chiecchio[1] and *Renzo Malvano*[2]

[1] Servizio di Fisica Sanitaria – Ospedale Mauriziano, Torino, Italy
[2] PROMM – Ospedale Mauriziano, Torino, Italy

Summary: As an alternative to the oversimplified error schemes currently adopted in establishing quality control (QC) strategies, a complex model was assumed implying

(a) the distribution of errors (critical error is regarded as a value discriminating between "effective errors" to be detected and "subcritical errors" which do not interfere with the medical decision whose detection is considered as a false-reject signal), and

(b) the possibility of simultaneous losses of precision and accuracy.

The control data recorded for digoxin radioimmunoassay over a one-year period were used for

(1) deriving the probability density functions of random and systematic errors, through a within-run across-level normalisation procedure;

(2) obtaining the functional relationships between the critical random or systematic error and the QC performance statistics (sensitivity, specificity, predictive value), weighted for the error prevalences, through integration of the probability density functions and the power functions associated with an exemplifying control rule; and

(3) describing the functions which correlate the corrected performance statistics with the allowable error (whose individual values account for all possible combinations of critical random errors and critical systematic errors), by extending to the tridimensional space the above procedures.

Analysis of the resulting data shows that it is necessary to revise the criteria for the choice and optimisation of QC schemes.

Introduction

The development of quality control (QC) strategies in clinical chemistry presupposes a statistically consistent evaluation of accept/reject criteria. Crucial contributions to this concept were provided by *Westgard*, *Groth* and coworkers (1–12) through the analysis of control rules, as characterised by the associated power functions obtained by simulation techniques (1, 3, 6), and their inclusion in the QC schemes (4, 7, 8, 10–12).

In the QC model of *Westgard* and *Groth*, the reference *stable analytical performance* is described by a random error inherent in the measurement procedure (possibly including both within-run and between-run components (2)) and by a constant method bias. The analytical errors to be controlled are considered as disturbances affecting the reference performance. In this approach, the error to be detected in order to maintain the quality within given specifications is defined as a *critical error* related to the *allowable analytical error*, which in turn refers to the medical decision limit (4, 6–8, 10). The simultaneous occurrence of random and systematic errors appears to be largely disregarded (1–11), though this possibility is considered in a paper concerning a QC program (12).

The QC strategy described in the literature essentially implies the binary classification *stable performance/critical errror*, ignoring other error situations (5, 7, 8, 10). In selecting a QC rule, the detection probabilities of reference error and of a given critical error are taken into account, using in some cases these data and the prevalence hypothesised for the critical error to define the positive predictive value of the rule itself (9, 10). No general criterion to establish such a prevalence is however provided.

A quite different model of "distributed error" may be proposed as an alternative. This is based on the most likely assumption that any cause of random or systematic disturbances could modify the reference situation to a variable and continuous extent, and that slight modifications should occur more frequently than the larger ones (in fact, the QC program referred to in l. c. (12) includes the case of distributed errors, but provides no guide to when and how such a distribution should be

considered). The critical error is then to be regarded as an error discriminating between "effective errors" to be detected and "subcritical errors", which do not interfere with the medical decision, and whose detection should be considered as a false-reject signal. (Note that in the proposed model, subcritical errors could either coincide with, or exceed, stable performance errors. Moreover, these latter errors are no longer considered as a reference).

The evaluation of error distributions and prevalences, their consequence in terms of QC performance, and the complications arising from the simultaneous losses of precisions and accuracy have all been taken into account in the present study, using as an experimental model the cumulative control data recorded for a digoxin radioimmunoassay. The results are presented and discussed.

Experimental Quality Control

The three-level QC results relative to a serum digoxin radioimmunoassay (Digoxin, Antibody Coated Tube – ^{125}I RIA Kit, ICN Pharmaceuticals, Costa Mesa CA, USA) were used, as obtained during a one-year period (138 successive runs, no exclusion). The overall variability (coefficient of variation, CV) ranged from 9.8% to 13.9%. Mean values and standard deviations (SD) were 0.50 ± 0.06 µg/l, 1.81 ± 0.19 µg/l and 3.16 ± 0.31 µg/l for the three control specimens, respectively.

These QC data should in fact correspond to a stable-performance state according to the *Westgard*'s approach. In the present study, however, the three control levels are not considered per se but they are intended as an example of a general situation, from which the distribution of both random and systematic errors might be evaluated, and a predictive error model might be derived, as explained in the following sections. To this purpose, a constant variability is assumed (in first approximation) for each concentration level, ignoring the differences actually found.

Computing Procedures and Results

Error distributions

A sequence of triplicate data was obtained by an across-level normalisation of the within-run QC results with respect to the overall mean values obtained for the three control specimens. Each triplicate set was regarded as being representative of individually defined assay conditions (daily analytical performance) assuming a uniform situation for the specimens, as stated above.

The standard deviation of each set of normalised triplicates was taken as the value of the observed imprecision. The deviations of triplicate means from the normalised general mean (target mean) were considered as the observed bias. For both imprecision and inaccuracy, prior densities for the expected errors were derived from total densities by empirical methods (13, 14), as described in Appendix A. The data obtained refer, of course, to within-run performance.

The resulting distribution histograms and the interpolating probability density functions are shown in figure 1, for both random and systematic errors. Arbitrary values in SD units were assigned to the error axis, assuming as unitary value the modal value of the random variability distribution (corresponding to a 5.8% CV). As expected, imprecision errors showed a right-skewed distribution (probability approaching zero for imprecisions approaching zero, maximum probability for SD = 1, lower probability for larger errors), while the bias errors were distributed symmetrically around the mean value (same probability for the occurrence of "negative" and "positive" deviations).

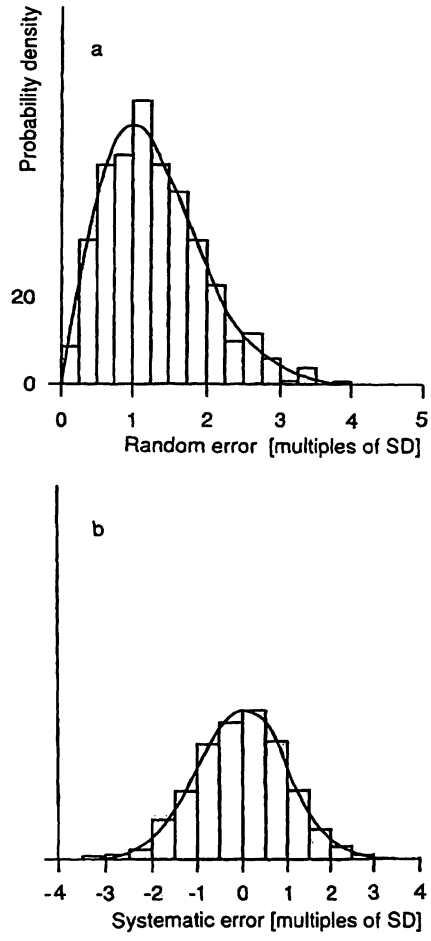

Fig. 1 Distribution frequency histograms and probability density functions related to the errors evaluated for a digoxin assay. The data were obtained from an across-level normalisation of the within-run quality control results relative to three control specimens, as explained in the text and in Appendix A. The modal value of random variability (5.8%) was arbitrarily taken as the unitary SD value. The multiples of SD are used to express both random variability (at the rop) and deviation from the target mean (at the bottom).

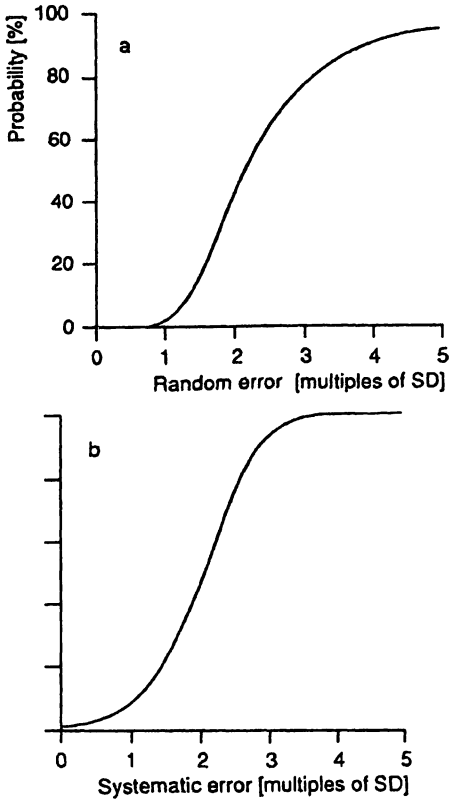

Fig. 2 Power functions associated with the control rule 1_{3s} (4 control observations). See text for the calculation procedure.

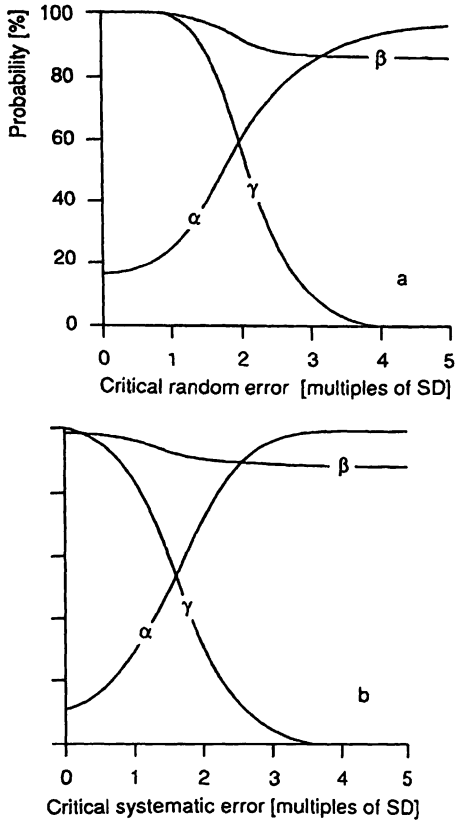

Fig. 3 Performance characteristics of the rule 1_{3s} (4 observations), weighted for the error prevalence, as a function of the critical error (at a 95% confidence limit). The data, relative to the model digoxin assay, were obtained from the power functions of figure 2 and the probability density functions of figure 1, following the calculation procedure reported in Appendix A. Also in this case, both errors are expressed as multiple of standard deviation. α: sensitivity, β: specificity, γ: predictive value of a reject.

It is assumed that errors pertaining to both a stable analytical performance (whatever it may be) and to its disturbances are described by these distributions.

Quality control rule and associated power functions

The QC rule 1_{3s} (the reject-signal implies that one control observation exceeds the target mean by 3 SD), for four observations, was used as an example. (A rule implying a number of observations different from three was purposely employed to avoid a misleading confusion between an operational step in establishing the experimental model and the rule itself). The related power functions for both random and systematic error were obtained by a computer simulation procedure following ther reference literature (1, 3) (10^5 extractions for each error value; random and systematic errors ranging 0 to 5 SD; 0.25 SD error increments in both cases). The functions are shown in figure 2.

Effects of error prevalence on the quality control performance

Each value assumed as critical error (either random or systematic error) on the x-axis identifies two distinct regions under the distribution curves of figure 1 and the power functions of figure 2, which correspond to subcritical and effective errors, respectively. In correspondence with any critical error, the reject probability, weighted for the prevalence of either effective erros (i. e, the "true positive" probability) or subcritical erros (i. e. the "false positive" probability), can be readily calculated by simple integration of the probability density functions and power functions (see details in Appendix B). A similar procedure can be followed for the accept probability ("true" or "false negative" probability, depending on whether subcritical or effective errors are concerned). Hence, as seen in figure 3, sensitivity, specificity and positive predictive values, all weighted for the error prevalences, may be obtained as a function of the critical error (random or systematic error). Weighting for prevalences implies a difference with respect to the standard definitions (*Bayes*ian model) of the terms "sensitivity" and "specificity", as adopted in this study. This unavoidably arises from the unusual situation generated by "distributed errors".

The following points should be considered:

– Sensitivity for effective error detection expectedly increases asymptotically with the size of critical error. (Note that in the hypothetical case of critical errors approaching zero, sensitivity is given by the integration of the whole areas under the probability density functions and the power functions; values of ca. 15% and 12%

thus result for random and systematic error, respectively).

– When critical errors approach zero, subcritical errors also obviously approach zero, so that false-reject signals are minimized, and specificity is maximal. After an initial decrease, the specificity towards subcritical errors tends to stabilise when slight increments in the subcritical error region correspond to slight increments in detection probability.

– When the critical error increases, the detection probability to effective errors is enhanced, but so is the risk of false-reject signals, due to the higher proportion of subcritical errors. The predictive value of reject therefore decreases, approaching null values as soon as the subcritical errors (and the related error signals) prevail over the effective errors.

Effects of the simultaneous occurrence of random and systematic variability on the quality control performance

The concepts elucidated above for imprecision and bias errors considered separately can be extended to the total error, by transposing the previously described bidimensional representations to the tridimensional space. Thus, instead of referring to given values for critical random or systematic variability related to the axis of abscissas (error axis), pairs of imprecision and bias values will now be involved, which individuate single points in the "error plane". A certain reject probability and a certain probability density can be associated with any such pair; in the tridimensional space, these will describe functional "power surface" and "density surfaces" (see the examples of fig. 4), exactly corresponding to the previous curves (power and density functions).

Also the concept of allowable analytical error may be extended to the error plane. In the new situation, the allowable error will individuate in this plane an infinite number of discriminating points, i. e. a "frontier" which discriminates the unacceptable combinations of critical random errors and critical systematic errors (combinations in the "reject" domain) from the acceptable ones (combinations in the "accept" domain). In figure 5 such a representation is schematically shown, while the computational procedure of the approach is given in Appendix C.

In correspondence with any value of allowable error, the frontier in the error plane is projected on the functional surface, i. e. power and probability density surfaces, as exemplified for the former in figure 6. Proceeding now to numerical integration of the power and density functions either in the reject or in the accept domain (the model of Appendix B still applies, though the calculations are somewhat more complicated), the data for sensitivity, specificity and predictive value may be derived for any given allowable error. An example is given in figure 7 for the same case (same rule, same error prevalences) already shown separately in figure 3 for imprecision and bias.

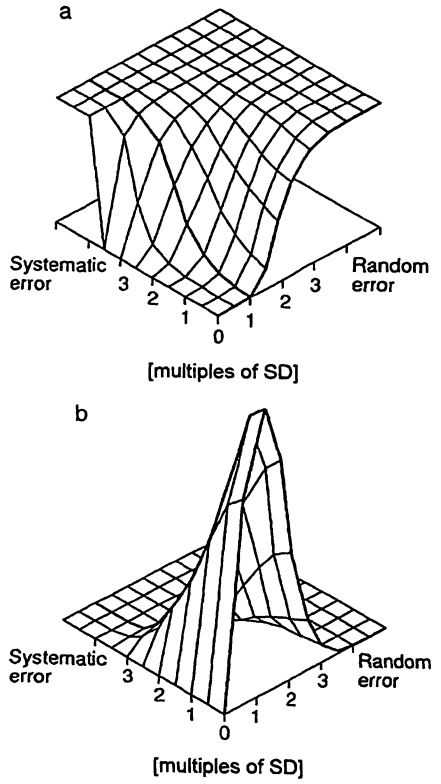

Fig. 4 Power functions for the rule 1_{3s} (4 observations) (a) and probability density functions (same cases as in fig. 1) (b) for the simultaneous occurrence of systematic and random errors.
Errors are expressed as multiples of standard deviation (see fig. 1). For any possible value of the allowable error, the associated combinations of critical random and systematic errors defines a line (power function or probability density function) in the space. The entire set of such lines individuates a surface (power surface, a; probability density surface, b). Thick lines indicate the same functions as in figures 1 and 2 (plane of random error = 1, plane of systematic error = 0).

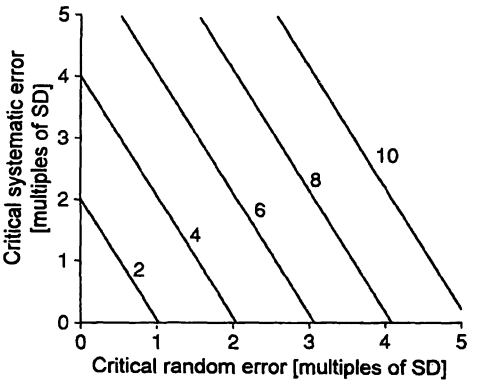

Fig. 5 "Frontier" lines identifying different values for the analytical allowable errors (arbitrary error units) in the critical error plane. Each of the functions indicated by a single value of allowable error (values 2 to 10 in the graph) represents all the possible combinations of critical random and systematic errors, according to the scheme given in Appendix C. The data refer to a 95% confidence level. It is worth noting that linearity of such frontiers confirms the situations found by other authors using different approaches for different applications (15, 16).

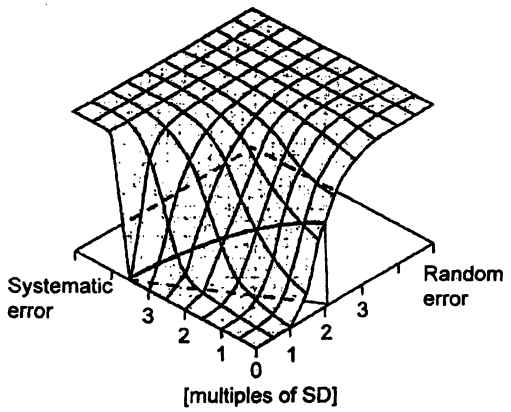

Fig. 6 As in figure 4a, exemplifying, in the case of power surfaces, the subdivision into two error domains separated by a frontier (surface regions corresponding to acceptable and unacceptable combinations of random and systematic errors). Errors are expressed as multiples of standard deviation. The dotted line indicates the frontier in the error plane (see fig. 5) corresponding to a given allowable error (equal to 3, in the example), while the thick line represents the projection of such a frontier on the functional surface. Assuming the numerical value of the allowable error (e. g. 3) and integrating both power functions (represented here) and probability density functions (omitted in the graph), the values for the associated performance statistics (sensitivity, specificity, positive predictive value for an allowable error = 3) can be obtained. Note that integrations (relative to a single-dimension error in Appendix B) here refer to a two-dimension error, making the calculations more complicated, but without affecting their scheme.

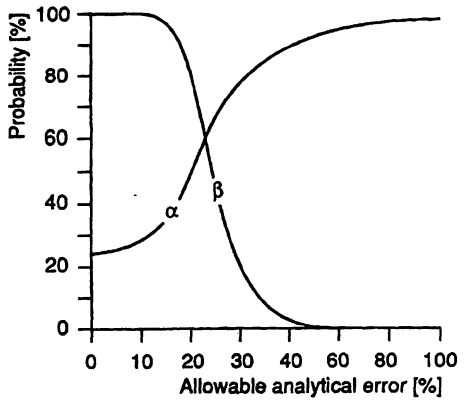

Fig. 7 Performance characteristics of the rule 1_{3s} (4 observations) weighted for the error prevalence, as a function of the allowable analytical error (at a 95% confidence limit). As compared to figure 3, the random and systematic errors are considered together. The allowable error is conventionally expressed as % (instead of multiples of SD as in the previous figures, unitary SD being taken as 5.8%). Since an evaluation of the control efficacy should proceed from the probability of effective error detection and the probability of true-reject signal, only the operational function sensitivity (α) and positive predictive value (β) are reported, omitting the specificity function.

In other words, reduction of the error plane to a single axis of error combinations (the allowable error axis) allows great simplification of data presentation (enabling the explicitation of bidimensional functions) and, hence, of the analysis of results.

It is worth noting that, under the conditions exemplified in figure 7, a satisfactory performance might not actually be attained in terms of sensitivity and predictive value, for any chosen analytical allowable error. As a matter of fact, at best a 60% probability is reached for both effective error detection and true reject signal for an allowable error of 23%, while the risks of false negatives and false positives dramatically increase when lower or higher allowable errors are involved.

Discussion

The "calibrated" control performance

A binary error model, presupposing oversimplified assumptions, implies an intrinsically simple QC strategy. Thus, the exemplifying 1_{3s} rule (for 4 observations) could be adopted (and an accept/reject decision could be made) by merely considering the associated power functions (see fig. 2), which indicate a specificity of about 98%, with respect to the reference conditions, and sensitivities of about 80%, or more, for both critical random errors exceeding 3 SD and critical systematic errors exceeding 2.5 SD.

The proposed model, in attempting to deal with the experimental complexity, instead generates greater difficulties. There is already evidence, for instance, that the above rule is inadequate in the specific case used as an example.

These few results appear to be sufficient to confirm the need for a revision of the criteria for choosing and optimising the QC procedures. Rather than evaluating the performance of the rules per se, preference should be given to a new concept of "calibrated" performance for any individual application, taking into account the actual error prevalences and the error acceptability limits.

In the attempt to calibrate the performance levels of QC, we could obviously use the whole set of control rules, their variants and their combinations, not excluding the AND schemes (where a positive outcome presupposes the violation of all the rules), as well as the usual OR combinations schemes (where a positive outcome implies that at last one of the rules concerned has been violated).

Moreover, it must be pointed out that the choice of the unitary error size (previously defined as "arbitrary") in fact represents a useful element of flexibility. As a matter of fact, in constructing a power function, any variation of the percentage error, when taken as the SD unit, is reflected by a modification of the control limits and, hence, of the control performance.

Examples of QC calibration, all referring to the model digoxin assay, are given in table 1 in relation to two hypothetical values (10% and 20%) chosen for the allowable analytical error. The effects of the manœuvres mentioned above are clearly illustrated by table 1. In particular, and with regard to the situation already seen in figure 2, a reduction of the control limits (e. g. a 60%

Tab. 1 Quality control performance as obtained with different schemes (data relative to the experimental model based on digoxin assay, assuming two different values for the allowable analytical errror, AAE)[a]

Quality control variables			Quality control performance			
No. of observations	Rule[b]	Variability corresponding to unitary SD[c] (%)	AAE = 10%		AAE = 20%	
			Sensitivity (%)	Predictive value of a reject (%)	Sensitivity (%)	Predictive value of a reject (%)
4	1_{3s}	5.8	19	100	57	75
4	1_{3s}	2.3	86	90	96	10
8	1_{3s}	5.8	44	100	82	72
8	1_{3s}	2.3	96	90	100	24
8	$1_{3s} + 2_{2s} + 3_{1s}$	5.8	89	94	99	40
8	$1_{3s} \cdot 2_{2s} \cdot 3_{1s}$	5.8	39	100	77	76

[a] In all instances, the procedure fully described in the text (for the case corresponding to the first-line condition of the table) was followed. Framed data indicate an acceptable quality control performance.

[b] The rules are indicated according to the usual notations derived from the literature (1–12). Rule combinations refer to OR(+) or AND (·) schemes.

[c] The modal CV value for imprecision (5.8%, fig. 1a) and the 40% of this value (2.3%) were used as an example.

decrease of unitary SD from 5.8% to 2.3%) appears sufficient to reach a satisfactory performance with respect to the lower allowable error. In this case, high performance levels are again obtained for single or combined (OR scheme) rules, by increasing the number of control observations (n = 8, which could, however, be economically inconvenient). A increased number of observations appears to be mandatory, in the case of the higher allowable error, in order to attain acceptable performance levels, using single or combined (AND) rules.

The predictive error model

The QC intended as a decisional tool presupposes the experimentally based probabilistic evaluations of a hypothesis. In this specific case, the hypothesis is that the presence of effective errors is actually indicated by the QC data, and that any error signal actually corresponds to the presence of effective errors. The definition of a predictive model for the errors therefore plays a crucial role in the QC decisional process.

The quantification of such a model involves some problems. First of all, a previous "history" of the test control must be available, as any prediction is founded on the extrapolation in time of historical information (and this obviously does not suit the case of newly adopted tests). It is then necessary to extract from the historical series of QC data the statistical material for the evaluation of random and systematic error distributions.

From this latter point of view, the across-level normalisation of the QC results relating to any individual run used in this study is not the only available procedure (note that substantially uniform behaviour towards the error sources at different concentrations, as implied by this approach, is hardly generalisable). Other procedures could, in principle, be based on within-run replicate estimates of the same control (replicates, however, are not generally available, or are limited to duplicates). Alternatively, they could involve contiguous sequences of QC data relative to limited time spans, for which a constant analytical performance might be reasonably supposed (this approach should either reduce the number of available data, or increase the duration of the data-collection period).

In any case, time and/or costs will be unavoidably involved in the preliminary phase needed to establish realistically consistent distributions of errors. Moreover, sophisticated mathematical procedures, and hence software packages not widely familiar to laboratory teams, should be introduced into routine QC.

Conclusions

In order to judge the actual role of QC in laboratory management, a series of questions should be answered.

Is QC to be considered as a decisional tool?

If so, is it necessary to establish in advance which errors may be tolerated and which may not?

Is it necessary to make a prediction for the occurrence of different-sized errors?

And should we expect, on the same occasions, to incur losses of both precision and accuracy?

If the answers are affirmative, then the procedures proposed by the authoritative researchers already quoted (1–13) should be regarded as necessary but insufficient steps, for providing the laboratory operators with effective and efficient decisional schemes.

Rather than indicating solutions, this paper is aimed at focusing attention on some crucial aspects of decision making in QC, and at indicating some tentative approaches. In particular, evidence was obtained for the critical nature of the evaluation of error prevalences and the selection of "calibrated control criteria", leaving aside, in this context, the problems set by the definition of allowable errors (see for instance l. c. (17)).

All these considerations contribute to the view that QC is a problem still open to further study, experience and discussion.

Acknowledgements

Thanks are due to Dr. *G. Vignati*, Ospedale Fornaroli, Magenta (Milan) who kindly supplied the quality control results of the digoxin assay, on which the experimental model was based.

This study was partially supported by Sclavo Diagnostics, Milan (Italy).

Appendix A

Empirical approaches to deriving the prior densities of expected errors from the total distributions of observed errors

The experimental distributions (*total distributions*) of both imprecision and inaccuracy errors can be represented as

$$h(re) = \int\int g(RE,SE)s(re|RE,SE)dREdSE, \qquad \text{Eq 1a}$$

$$h(se) = \int\int g(RE,SE)s(se|RE,SE)dREdSE, \qquad \text{Eq 2a}$$

where

re and se indicate the *observed errors* (random and systematic, respectively),

RE and SE indicate the *expected errors* (random and systematic, respectively),

g(RE,SE) is the a priori density of expected errors, and

s(re|RE,SE), s(se|RE,SE) are the distribution densities of the observed errors.

Owing to the statistical independence of imprecision and inaccuracy, Eqs 1a and 1b can be rewritten as

$$h(re) = \int\int g(RE)g(SE)s(re|RE,SE)dREdSE, \qquad \text{Eq 1b}$$

$$h(se) = \int\int g(RE)g(SE)s(se|RE,SE)dREdSE. \qquad \text{Eq 2b}$$

In the case of Eq 1b a simplification is possible since

$$s(re|RE,SE) = s(re|RE)$$

(as the distribution h(re) has been derived on the reasonable assumption that the three realisations are characterised by the same inaccuracy). Moreover,

$$s(re|RE) = \propto \frac{re}{RE} e^{-\left(\frac{re}{RE}\right)^2},$$

due to the kind of sampling error starting from three realisations.

Hence the following integral equation can be written

$$h(re) = \propto \int g(SE)dSE \cdot \int g(RE)\frac{re}{RE}e^{-\left(\frac{re}{RE}\right)^2} dRE$$

$$= \propto \int g(RE)\frac{re}{RE}e^{-\left(\frac{re}{RE}\right)^2} dRE \qquad \text{Eq 1c}$$

which can be resolved by numerical methods to find g(RE), starting from the experimental data relative to h(re) (14).

Once g(RE) has been obtained, accounting for the influence of random error RE on distribution s(se|RE,SE), and assuming for this latter a *Gaussian* shape (centered on 0), Eq 2b can be expressed as

$$h(se) = \propto \int\int g(RE)g(SE)e^{-\frac{1}{2}\frac{se^2}{(RE^2 + SE^2)}} \frac{dREdSE}{\sqrt{(RE^2 + SE^2)}}, \qquad \text{Eq 2c}$$

which can be resolved by numerical methods to find g(SE), starting from the experimental data relative to h(se) (14).

Appendix B

Calculation of the quality control performance characteristics, weighting by error prevalence

In the presence of effective errors (i. e. errors exceeding the critical error, E_C), true-reject signals (true positives, TP) and false-reject signals (false positives, FP) can be expressed as

$$TP(E_C) = \int_{E_C}^{\infty} u(E)g(E)dE,$$

$$FP(E_C) = \int_{0}^{E_C} u(e)g(E)dE,$$

where u(E) and g(E) indicate the power function of the chosen rule and the probability density of the error (random or systematic error), respectively.

True-accept signals (true negatives, TN) and false-accept signals (false negatives, FN) are given by the relationships

$$FN(E_C) = \int_{E_C}^{\infty} [1 - u(E)\, g(E)dE,$$

$$TN(E_C) = \int_{0}^{E_C} [1 - u(e)\, g(E)dE,$$

The prevalence of effective errors (f) can be described as

$$f(E) = \int_{E_C}^{\infty} g(E)dE = TP(E_C) + FN(E_C)$$
$$= 1 - TN(E_C) - FP(E_C).$$

Hence the QC performance characteristics weighted for the error prevalence can be derived:

$$\text{sensitivity}(E_C) = TP(E_C) / [TP(E_C) + FN(E_C)] = TP(E_C / f(E_C),$$

$$\text{specificity}(E_C) = TN(E_C) / [TN(E_C) + FP(E_C)]$$
$$= TN(E_C) / [1 - f(E_C)],$$

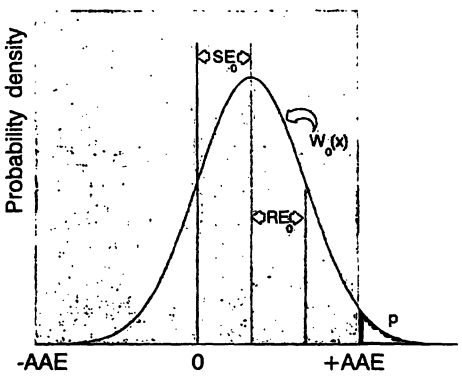

Fig. 8 Combination of random and systematic error. The allowable analytical error (AAE) is indicated by the shadowed region. The expression $w_0(x) = w(x;RE_0,SE_0)$ represents a generic *Gaussian* distribution of analytical estimates x for a single specimen, corresponding to a given critical combination RE_0, SE_0 $=(RE,SE)_{0,C}$ of errors (RE, random error; SE, systematic error); p is the fraction of the distribution, excluded at a sought confidence level. The representation follows the scheme reported for critical errors RE and SE separately considered (7, 8, 10).

positive predictive value (E_C) = $TP(E_C) / [TP(E_C) + FP(E_C)]$

$$= TP(E_C) / \int_0^\infty u(E)g(E)dE \propto TP(E_C).$$

Appendix C

Calculation of the frontier on the error plane

Considering figure 8, the definition of the frontier will be:

$$\text{frontier}(AAE, p) = \langle (RE,SE)_C | 1 - p = \int_{-\infty}^{AAE} w(x|RE,SE)dx \rangle,$$

i.e. the set of the critical combinations $(RE,SE)_C$, so that $1 - p = \int_{-\infty}^{AAE} w(x|RE,SE)dx$, where $w(x|RE,SE)$ refers to the distributions of replicated measurements, exemplified in figure 8.

References

1. Westgard JO, Groth T, Aronsson T, Falk H, de Verdier CH. Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. Clin Chem 1977; 23:1857–67.
2. Westgard JO, Fak H, Groth T. Influence of between-run component of variance, choice of control limits, and shape of error distribution on the performance characteristics of rules for internal quality control. Clin Chem 1979; 25:394–400.
3. Westgard JO, Groth T. Power functions for statistical control rules. Clin Chem 1979; 25:863–9.
4. Westgard JO, Groth T. Design of statistical control procedures utilizing quality specifications and power functions for control rules. In: Wilson DN, Gaskell SJ, Kamp KW, editors. Quality control in clinical endocrinology. Cardiff: Alpha Omega Publ 1979:155–64.
5. Westgard JO, Groth T. A multi-rule Shewart chart for quality control in clinical chemistry. Clin Chem 1981; 27:493–501.
6. Westgard JO, Groth T. Design and evaluation of statistical control procedures: applications of a computer "quality control simulator" program. Clin Chem 1981; 27:1536–45.
7. Groth T, Falk H, Westgard JO. An interactive computer simulation program for the design of statistical control procedures in clinical chemistry. Computer Programs Biomed 1981; 13:73–86.
8. de Verdier CH, Groth T, Westgard JO. What is the quality of quality control procedures? Scand J Clin Lab Invest 1981; 41:1–14.
9. Westgard JO, Groth T, de Verdier CH, A predictive value model for quality control effects of prevalence of errors on the performances of control procedures. Am J Clin Pathol 1984; 80:49–56.
10. Westgard JO, Groth T, de Verdier CH. Principles for developing improved quality control procedures. Scand J Clin Lab Invest 1984; 44 Suppl 172:19–41.
11. Aronsson T, Groth T. Nested control procedures for internal analytical quality control. Theoretical design and practical evaluation. Scand J Clin Lab Invest 1984; 44 Suppl 172:51–64.
12. Groth T, Falk H, Westgard JO. A quality control simulator for design and evaluation of internal quality control procedures. Scand J Clin Lab Invest 1984; 44 Suppl 172:195–201.
13. Feller W. An introduction to probability theory and its application. 3rd ed, vol 1. New York: J Wiley & Sons 1968:266–70.
14. Tikhonov A, Arsénine V. Méthods de résolution de problèmes mal posés. Traduction française. Moscou: Edition Mir 1976:116–34.
15. Westgard JO, Hyltoft P, Wiebe DA. Laboratory process specifications for assuring quality in the US National Cholesterol Education Program. Clin Chem 1991; 37:656–61.
16. Westgard JO. Charts of operational process specifications ("OPSpecs charts") for assessing the precision, accuracy, and quality control needed to satisfy proficiency testing performance criteria. Clin Chem 1992; 38:1226–33.
17. Fraser CG. Generation and application of analytical goals in laboratory medicine. Ann Ist Super Sanità 1991; 27:369–76.

Received May 8/October 4/December 6, 1995

Corresponding author: Dr R. Malvano, Servizio di Fisica Sanitaria – Ospedale Mauriziano, Largo Turati, 62, I-10128 Torino, Italy